

Claims

What is claimed is:

1 1. A method of processing audio-based data associated with a particular language,
2 the method comprising the steps of:

3 storing the audio-based data;

4 generating a textual representation of the audio-based data, the textual
5 representation being in the form of one or more semantic units corresponding to the audio-
6 based data; and

7 indexing the one or more semantic units and storing the one or more indexed
8 semantic units for use in searching the stored audio-based data in response to a user query.

1 2. The method of claim 1, wherein the semantic unit is a syllable.

1 3. The method of claim 2, wherein the syllable is a phonetically based syllable.

1 4. The method of claim 1, wherein the semantic unit is a morpheme.

1 5. The method of claim 1, wherein the generating step comprises decoding the audio-
2 based data in accordance with a speech recognition system.

1 6. The method of claim 5, wherein the speech recognition system employs a
2 semantic unit based language model.

1 7. The method of claim 1, wherein the indexing step comprises time stamping the
2 one or more semantic units.

1 8. The method of claim 1, wherein the searching step comprises:

2 processing the user query to generate one or more semantic units representing the
3 information that the user seeks to retrieve;

4 searching the one or more indexed semantic units to find a substantial match with
5 the one or more semantic units associated with the user query; and

6 retrieving one or more segments of the audio-based data using the one or more
7 indexed semantic units that match the one or more semantic units associated with the user
8 query.

1 9. The method of claim 8, wherein the searching step further comprises presenting
2 the retrieved data to the user.

1 10. The method of claim 1, wherein the particular language is an Asian based
2 language.

1 11. The method of claim 10, wherein the particular language is Chinese.

1 12. The method of claim 11, wherein the semantic unit is a Chinese character.

1 13. The method of claim 1, wherein the particular language is a Slavic based
2 language.

1 14. The method of claim 1, wherein the one or more semantic units are indexed
2 according to speaker attributes.

1 15. The method of claim 1, wherein the one or more semantic units are indexed
2 according to at least one of when the audio based data was produced and where the audio
3 based data was produced.

1 16. The method of claim 1, further comprising the step of storing video based data
2 associated with the audio based data for use in searching the stored audio based data and
3 the video based data in response to a user query.

1 17. The method of claim 16, wherein the searching step includes a hierarchical
2 search routine.

1 18. The method of claim 1, wherein the generating step comprises stenographically
2 transcribing the audio-based data to generate the textual representation.

1 19. Apparatus for processing audio-based data associated with a particular
2 language, the apparatus comprising:

3 at least one processor operative to: (i)store the audio-based data; (ii) generate a
4 textual representation of the audio-based data, the textual representation being in the form
5 of one or more semantic units corresponding to the audio-based data; and (iii) index the
6 one or more semantic units and store the one or more indexed semantic units for use in
7 searching the stored audio-based data in response to a user query.

1 20. An audio-based data indexing and retrieval system for processing audio-based
2 data associated with a particular language, the system comprising:

3 memory for storing the audio-based data;
4 a semantic unit based speech recognition system for generating a textual
5 representation of the audio-based data, the textual representation being in the form of one
6 or more semantic units corresponding to the audio-based data;
7 an indexing and storage module, operatively coupled to the semantic unit based
8 speech recognition system and the memory, for indexing the one or more semantic units
9 and storing the one or more indexed semantic units; and

10 a search engine, operatively coupled to the indexing and storage module and the
11 memory, for searching the one or more indexed semantic units for a match with one or more
12 semantic units associated with a user query, and for retrieving the stored audio based data
13 based on the one or more indexed semantic units.

09437971-11099